From: *Restoring Trust in Organizations and Leaders: Enduring Challenges and Emerging Answers*, Roderick M. Kramer and Todd L. Pittinsky (eds.), New York: Oxford University Press, 2012, pp. 151-169.

7 The Elasticity of Trust

How to Promote Trust in the Arab Middle East and the United States

IRIS BOHNET, BENEDIKT HERRMANN, MOHAMAD AL-ISSISS, ANDREA ROBBETT, KHALID AL-YAHYA, AND RICHARD ZECKHAUSER

INTRODUCTION

To trust is to risk. When we lend someone money, we make ourselves vulnerable, hoping or expecting that the borrower will reward our trust and return the money at a later stage, possibly with interest or a reciprocal favor added. Generally, people are more willing to engage in a risky activity, such as buying a stock or starting a business, the greater the expected returns from the activity. This chapter examines whether willingness to trust follows the same logic—that is, whether it responds to changes in the expected value of trusting, much like willingness to take risk responds to changes in the expected value of risk taking.

We refer to the responsiveness of trust to changes in the expected value of trusting as the elasticity of trust. A better understanding of how elastic trust is will help us choose the most effective means to promote trust. The effectiveness of an intervention designed to promote trust, such as introducing insurance or legal remedies in case of a betrayal, will depend on the context and culture in which enhanced trust is sought. To demonstrate this point, we look at trust behavior and its response to two interventions in three countries of the Arab Middle East (Jordan, Saudi Arabia, and the United Arab Emirates [UAE]) and compare our findings with the context in which the preponderance of trust research has taken place so far, the United States.

We employ a binary-choice trust game to measure willingness to trust. Thus, a first mover decides whether or not to trust a (typically anonymous) counterpart, the second mover (Camerer & Weigelt, 1988; Dasgupta, 1988; Kreps, 1990). (For expository convenience, we refer to the first mover with female and the second mover with male pronouns. Note, however, that all our subjects were anonymous and thus gender was not a treatment variable.) If the first mover does not trust, she is said to play "sure"—that is, not to trust. The game ends and both parties receive a moderate payoff, [M,M]. If she trusts, the second mover can either reward or betray trust. If he rewards trust, both he and the first mover are better off than if trust had not been offered, thus yielding payoff [B,B]. If the second mover betrays trust, he receives the highest possible payoff and the first mover the lowest possible payoff [L,H]. Thus H>B>M>L. Trust increases the second mover's payoff whatever

he chooses, and the total payoff to the two parties. We also posit, as in many trust games, that (H + L) = (B + B).

As H>B, the second mover has a financial incentive to betray, making the first mover worse off than if she had not trusted, since L<M. The ordering of payoffs implies that a money-maximizing second mover prefers to betray rather than reward trust, and that a money-maximizing first mover prefers not to offer trust if she expects to be betrayed. In this sequential game, the unique Nash equilibrium has the first movers always play sure, which terminates the game.

The first mover's expected value from trusting can be affected by increasing B, the better payoff she and her counterpart receive if her trust is rewarded and/or by increasing L, the low payoff she receives if her trust is betrayed. Similarly, the more optimistic is the first mover about the second mover's likelihood of trustworthiness, the more she will trust. We will examine how elastic trust is to changes in the payoff from trusting as well as to changes in the likelihood of trustworthiness.

We explore an anonymous one-shot interaction. In such an environment, if players were merely maximizing their own expected payoffs, second movers would never reward trust, implying that first movers would never trust. Other factors may enter, however, and thus affect the elasticity of trust. We are particularly interested in how social preferences, both of the first and the second mover, influence how responsive trust is to changes in the value of trusting and the likelihood of trustworthiness. Economists define social preferences as non-selfish, other-regarding motives of behavior. Social preferences may be based on outcomes only—for example, with people caring about the relative share of the pie they receive compared to relevant others ("inequality aversion"; e.g., Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999, 2002; Loewenstein et al., 1989).

Social preferences can also depend on intentions, taking into account how outcomes came to be (e.g., Blount, 1995; Rabin, 1993). Beliefs about causation have been found to have dramatic effects on behavior in a variety of contexts (e.g., in social dilemma situations, see Rutte et al., 1987, and in bargaining contexts, see Pillutla & Murnighan, 1996). In a trust interaction, intention-based social preferences may lead second movers to reward trust because they feel a responsibility to reciprocate the first mover's kindness. Similarly, first movers may care about how the final outcome comes to be and may not trust because they do not want to expose themselves to the possibility of betrayal by another human being. People are said to be betrayal-averse if they dislike being betrayed in addition to its negative material consequences (Bohnet et al., 2008). Individuals who are betrayal-averse are less willing to take the same risk with the same payoffs when the agent of uncertainty is another person rather than nature (e.g., a chance device).

All of these social preference models assume that people judge behavior relative to some reference point—for example, the share of the pie a reference group receives in inequality aversion models, or some norm of acceptable behavior in the case of reciprocity or betrayal aversion models. Reference points may vary across cultural contexts. In fact, building on Bohnet et al. (2010), we argue that one important role of culture is to define reference points across various game-theoretic contexts, such as bargaining or cooperation situations. For example, Henrich et al. (2010) showed that fairness norms were related to the degree of

market integration in a given society. Reference-dependent models allow us to understand why certain behaviors are unacceptable or even punished in some societies but not in others. Reference points shape expectations and deviations from them lead to disappointment and a sense of loss, much in the manner of traditional prospect theoretic models of loss aversion where outcomes are evaluated relative to a reference point and either perceived as gains or as losses (Kahneman & Tversky, 1979; Köszegi & Rabin, 2006).

Reference points for trustworthiness differ between the countries studied in the Arab Middle East and the United States. In the tribal societies of the Middle East, trust interactions overwhelmingly take place within groups. In groups, repeated game incentives and reputational concerns motivate people to be trustworthy. In the United States, by contrast, there are many trust interactions among strangers, with formal institutions, such as contract law, making trust possible. The relation-based trust of the Middle East focuses on preventive mechanisms that eliminate or at least decrease the likelihood of betrayal; the rule-based trust of the United States focuses on mitigating the costs of betrayal, for example by offering damages to the insured party.

The prevention approach of the Arab Middle East leads to higher reference points of trustworthiness than does the mitigation of the United States. If the elasticity of trust depends on reference points, trust will be very inelastic until one's reference point is reached. This implies that the overall elasticity will be lower in prevention- than in mitigation-based regimes—that is, in the Arab Middle East than in the United States.

The chapter is organized as follows. First we explore cross-societal factors that might affect how elastic trust is to changes in both the value of trusting and the likelihood of trustworthiness. Then we present evidence on how mechanisms affecting the value of trusting, namely a decrease in the cost incurred by the betrayed party, such as damages, affect the elasticity of trust. We then examine the impact of mechanisms decreasing the likelihood that a violation of trust will occur.

■ PREVENTION IN THE ARAB MIDDLE EAST AND MITIGATION IN THE UNITED STATES

We present two observations about important differences between the three countries of interest in the Arab Middle East and the United States that might affect the elasticity of trust without making any claims about their relative importance. Many of our observations about the Arab Middle East are based on the Arab Human Development Reports (AHDR, United Nations Development Program, 2002–04), which are among the few sources that provide comparable data on Arab countries. They were written by Arab scholars.[1] This part extends our earlier work on trust in

1. The reports emphasize that all its authors are Arab, and thus provide an insider's look at the conditions in the Arab world. The reports were warmly received in the West (e.g., hailed by the *New York Times*, the *Financial Times*, and *The Economist*) but received mixed reviews in the Arab world. Criticism centered around the reports' recommendations and conclusions and not around their description of

the Arab Middle East, where we focused on Kuwait, Oman, and the UAE (Bohnet et al., 2010).

Observation 1: Western Law and Islamic Law

In the United States, various arrangements that decrease the costs of betrayal help to encourage trust. Damages for betrayal are part of nearly all contractual arrangements. Contract law also allows for "efficient breach," where the first mover is paid her loss by the second mover should the latter find it beneficial to breach the contract. "Perfect expectation damages," for example, compensate the first mover for any injury caused by the second mover. They make the potential victim of breach equally well off financially whether the contract is performed or breached. In fact, Oliver Wendell Holmes Jr., the famed American jurist, wrote about U.S. law in 1897: "The duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it and nothing else" (Rosen, 2000, p. 139). Even where the legal system falls short, for example, when transaction costs are large, commercial insurance may protect a contracting party for losses ranging from breach of contract (surety bonds) to employee theft (fidelity bonding).

Such contractual arrangements are rarely found in Islamic law, which plays an important role in many countries of the region, including the three that are our focus here: Jordan, Saudi Arabia, and the UAE.[2] The availability of compensation would justify and encourage taking risk (holding other factors constant), which could be undesirable as certain forms of risk taking, "gharar," are prohibited in Islamic law. Sami Ibrahim Al-Suwailem, an Islamic economist[3] from Saudi Arabia, explains: "prohibition of *gharar* is established on the general principle that a decision maker shall not rely on pure chance to achieve desired outcomes. The approach is suitable not only for personal decisions, but also for interactions with others. It is a principle that governs general human behavior under risk" (2000a, p. 9). Specifically regarding damages, he writes: "whatever costs the agent incurs can be compensated for only in case of success; in case of failure, they represent a source of sorrow and regret" (Al-Suwailem, 2000a, p. 6).

Incomplete contracts based on trust and trustworthiness are inherently risky. They are referred to as "gharar contracts." "Gharar contracts are often dynamically

the social, political, and economic environment, on which we will focus here. For a discussion of the comments on the reports, see Baroudi (2004).

2. Scholars of Islamic law agree that the shari`ah is not just law in the usual Western sense; "Rather, it is a divine command governing all human behavior, whether concerned with this world or the next. [p. 10] . . . the law never accedes to any principle that worldly enforcement applies only to matters of 'law' rather than to 'morals'. . . Hence, nothing corresponds consistently to the modern Western line of division between religious (or moral) and legal" (p. 16) (Vogel, 1997).

3. See Kuran (1995, 2004) for a definition of Islamic economics and a discussion of its various proponents. He writes in the 1995 article (p. 159): "its agents act under the guidance of norms drawn from the traditional sources of Islam. . . These norms 'command good' and 'forbid evil'. . . The intended effect of the norms is to transform selfish and acquisitive *homo economicus* into a paragon of virtue, *homo Islamicus*. *Homo Islamicus* acquires property freely, but never through speculation, gambling, hoarding, or destructive competition."

inconsistent, and therefore it is often not in the best interest of one of the parties to fulfill the contract. . . . Non-gharar contracts in contrast can be fulfilled by self-interest of involved parties" (Al-Suwailem, 2000b, p. 95). Observation 1 supports the notion that the mitigation of cost has the ability to play a substantially larger role in fostering trust in the United States than it does in the Arab Middle East.

Observation 2: Societal Organization

The United States can be characterized as an "individualist" country. By contrast, the Arab Middle Eastern countries under study—Jordan, Saudi Arabia, and the UAE—are "collectivist" (e.g., Hofstede, 2001; Triandis, 1995; Inglehart et al., 2005). Such cultural theory assessments predict greater "uncertainty avoidance" and a stronger distinction between "ingroup" and "outgroup" members in the latter than the former. The Arab Human Development Reports stress the predominance of group-based societal organization in the Arab world. "Clannism (al-'asabiya), in all its forms (tribal, clan-based, communal, and ethnic), tightly shackles its followers through the power of the authoritarian patriarchal system. This phenomenon . . . represents a two-way street in which obedience and loyalty are offered in return for protection, sponsorship, and a share of the spoils. . . Its positive aspects include a sense of belonging to a community and the desire to put its interests first" (AHDR, 2004, p. 145). Hisham Sharabi (1988), a leading Palestinian intellectual, describes the Arab world as a "neopatriarchy," where social relations are more vertical than horizontal (as in the typical modern Western state) and where society is stratified according to family and clan membership.

Building on the data available in the Arab world from the World Values Survey (2003), the AHDR (2004) shows that confidence in political institutions is very low in both Jordan and Saudi Arabia. (The UAE was not part of the survey.) It concludes that "clannism flourishes . . . wherever civil or political institutions that protect rights and freedoms are weak or absent. Without institutional supports, individuals are driven to seek refuge in narrowly based loyalties that provide security and protection" (AHDR, 2004, p. 146). A group-based societal organization based on longstanding relationships can substantially reduce the social uncertainty involved in trust. Within groups, repeated interactions are likely, information on reputation spreads quickly, monitoring is comparatively cheap, social sanctions help maintain commitments, and loyalty brings high levels of reliability. Disloyalty is often punished by expulsion (e.g., Layne, 1987). Observation 2 suggests that the prevention of betrayal plays a substantially more important role in fostering trust in the Arab Middle East than in the United States.[4]

4. Building on these insights, Greif (1993, 1994) shows how in a collectivist equilibrium, first movers spend resources to acquire information on a second mover's past behavior and deem a second mover trustworthy only if he has rewarded trust in the past. In contrast, in an individualist equilibrium, first movers do not care about a second mover's history but induce trustworthiness by compensating second movers with higher wages than paid in the collectivist equilibrium. His case studies for collectivist arrangements are the Maghribi traders of the 11th century, Jews who were part of the Muslim world, and for individualist solutions the Genovese traders of the 12th century, who were part of the Christian world.

To get a better sense of the importance of group-based trust in our study population, we administered a short post-experimental questionnaire in the United States and the UAE to ask our participants what fractions of people of their own kind and various other groups (e.g., Africans, Americans, Emiratis, Europeans) they would be willing to trust. The questionnaire appears as Table 7-A1 in the Appendix. We recognize the shortcomings of questionnaires. For example, political correctness may play a different role in the United States than in the UAE, possibly leading to higher reported levels of trust for members of groups not really trusted in the former than the latter country.

We report summary data here that should be interpreted with the caveats mentioned above in mind. The difference in the fraction of people trusting ingroup members (Emiratis in the UAE and Americans in the United States) and the fraction trusting outgroup members (all others) normalized by the fraction of people trusting an ingroup member is 0.42 in the UAE and 0.09 in the United States. That number would be 1 if people trusted some or all ingroup members but no outgroup members, zero if the fraction trusting ingroup members was identical to the fraction trusting outgroup members, and negative if people preferred to trust the outgroup to the ingroup. More generally, including other categories such as religion and gender, ingroup/outgroup distinctions are much more pronounced in the UAE than in the United States, with an ingroup preference for people with the same religion of 0.75 in the UAE and 0.01 in the United States (compared to all other major religions) and a same-sex preference of 0.31 for Emirati men and −0.07 for American men. (That is, American men trust women more than men.) Neither Emirati nor American women report any sex-based distinctions.

The cross-regional differences in the legal and societal structures suggest that trust is produced differently in the three Arab Middle Eastern countries under study than in the United States. Formal institutions help foster trust in the United States. Prime responses to the risk of betrayal are the nature of contracts, which include damages for breach, and allow for various insurance arrangements. These instruments allow people to reduce the expected costs of betrayal, thereby turning trust into principally an investment decision. They foster trust between strangers and allow for trust across groups. The parties do not have to know their counterpart's type or reputation before offering a contract; the legal system protects them against material losses if they are betrayed. When Americans engage in business transactions, terms such as "honor," "betrayal," and "loyalty" do not play a predominant role, at least in recent times. It is more calculation than principle or emotion that leads to the decision to trust.[5]

5. Recent work by Hsee and Rottenstreich (2004) looks at the difference between calculation-based behavior and emotion-based behavior in the context of decisions by unitary actors. They find that with willingness-to-pay questions, calculation-based behavior is much more responsive to (elastic to) quantitative magnitudes (e.g., number of CDs received or number of animals saved). They show that individuals can be primed to engage in calculation-based behaviors by giving them specific questions asking them to do calculations.

By contrast, loyalty, reputation, and the fear of adverse treatment from the group are the main factors producing trust in the Arab Middle East. The response to disloyalty is exit or expulsion. Emiratis, Jordanians, and Saudis "exit" from a relationship and do not trust unless loyalty or trustworthiness is virtually guaranteed. Bohnet et al. (2010) argued that such differences in trust production lead to different accustomed levels of trustworthiness in the two regions. "Prevention regimes," by definition, focus on preventing the occurrence of betrayal. Thus, people experience trustworthiness more often than people in "mitigation regimes," affecting what people expect appropriate trustworthiness levels to be and how elastic trust is to changes in the expected value of trusting.

■ TRUST RESPONSIVENESS TO CHANGES IN THE VALUE OF TRUST

The value of trusting can be enhanced by increasing the benefits of a successful trust interaction and by decreasing the cost of one that is unsuccessful. We focus on the latter, on mechanisms that decrease the material costs of betrayal, such as damages in case of breach and insurance provisions. Such measures decrease the risk involved in trusting, holding everything else constant. Much like fire insurance allows people to take the chance of building a factory that might burn to the ground, so too we might expect damages for betrayal to enable people to trust strangers. However, in contrast to nature-made catastrophes, the agent of uncertainty in trust interactions is one or more other people who might respond to whatever changes to the expected value of trusting we introduce. "Holding everything else constant," as suggested above, is not a viable assumption in a trust interaction. Once damage or insurance provisions are put in place, the behavior of the parties involved, first and second movers, might change.

The logic on how behavior changes is easiest to see if we start with the second mover. We consider three motivations: (1) self-interest, (2) inequality aversion, and (3) reciprocity. Focusing on the second mover, a self-interested, money-maximizing second mover is not affected by changes in the first mover's betrayal payoffs; all he cares about are his own payoffs. Allowing for inequality aversion, however, changes the game. Decreasing the losses incurred by the betrayed party also diminishes the payoff differences between the betrayer and the betrayee. The betrayer still receives the superior payment, but the betrayed player is not as bad off as she would have been absent compensation for losses. In the trust game introduced earlier, with a first mover offering trust and a second mover deciding whether or not to reward it, rewarding yields equal payoffs for both and betraying yields higher payoffs for the second than for the first mover, who ends up worse off than if she had never trusted. A second mover who is inequality averse will have conflicting incentives. Betraying will increase his own payoff, which is good, but will also increase the gap between his payoff and that of the first mover, which is bad. Thus, holding his own benefit from betrayal fixed, the smaller is that gap (holding payoffs from trust-reward fixed), the more likely he is to betray.

Finally, a second mover's willingness to reward trust may also be affected by intention-based preferences. In reciprocity models (e.g., Rabin, 1993), a second

mover is more likely to reciprocate a first mover's kindness, the kinder he perceives her behavior to be (compared to some reference level). Thus, a reciprocation-oriented second mover, one who takes the first mover's payoffs in case of betrayal as a signal of how vulnerable she is willing to be in a given interaction, will be less likely to reward trust, the less risk the first mover is taking—that is, the higher are her payoffs in case of betrayal.

Consider merely the first mover's betrayal payoff, holding all other payoffs fixed. Both, inequality aversion and reciprocity, will lead second movers to be more inclined to reward trust, the smaller is the first mover's betrayal payoff. If first movers expect second movers to have such inclinations, they should be more likely to trust, the smaller are their own betrayal payoffs. Thus, the first mover confronts a difficult optimization problem: a decrease in her losses in case of betrayal enhances her value from trusting, but also makes it more likely that she will be betrayed.

The effect on the first mover thus is unclear. If she focuses on the change of the value of trusting without considering the effect such a payoff change might have on the second mover's willingness to reward trust, she would be more likely to trust when damages are offered. If she takes both possible effects into account, they will counterbalance each other, making the effect on her behavior ambiguous. If she focuses only on the second mover's response, she might even trust less if her betrayal payoffs are higher. For second movers, the prediction based on these two social preferences is clear: other factors equal, they should reward trust less, the higher is the first mover's betrayal payoff.

Moving on to the first mover, allowing for inequality aversion does not further complicate matters. Both inequality-averse first movers, those who care about the payoff difference between themselves and the second mover, as well as selfish first movers, those who care only about their own payoffs, would be more likely to trust the larger is their own betrayal payoff.

A further complication arises, however, if we allow for first movers to be affected as well by betrayal aversion (Bohnet et al., 2008). Betrayal aversion is a separate cost. It diminishes the relative importance of interventions that decrease the material cost of betrayal, such as damages or insurance. In the extreme, if concerns about betrayal are extremely large relative to concerns about the monetary consequences of an action, trust will be completely inelastic to any changes in the material cost of betrayal.

Mechanisms that decrease the material cost of betrayal, such as damages and insurance, are more common in Western than in Arab Middle Eastern countries. According to Vogel (1997) and Al-Suweilam (2000a,b), the conservative use of mitigation measures in the Arab Middle East is not just a descriptive but also a prescriptive norm, suggesting that such mechanisms should not be relied on in trust interactions. Accordingly, we expect people to be less responsive to changes in the material cost of betrayal in the Arab Middle East than in the West.

Al-Ississ and Bohnet (2011) examined this conjecture experimentally, employing a standard binary-choice trust game in Jordan and the United States. They focused on the most basic accomplishment of insurance and damages,

namely that they decrease the losses incurred by the betrayed party. They varied the low betrayal payoff, L, from 10% of the moderate payment, M, a first mover would earn if she had not trusted, to 90% of that payoff, with other payments held constant. More precisely, in a between-subject design, experimental participants were confronted with one of five possible betrayal payoffs, L, for the first mover: 0.1M, implying basically no insurance in case of betrayal, 0.3M, 0.5M, 0.7M, or 0.9M. (The last one implies almost full insurance in case of betrayal.) Thus, by varying the material cost of betrayal, they measured how first movers' willingness to trust—that is, choosing trust rather than playing safe—and the second mover's willingness to reward trust—that is, choosing reward rather than betraying—was affected by different degrees of insurance. Note that changing the first mover's payoff in case of betrayal also affects overall efficiency.

In both countries, they found that as the first mover's payoff given betrayal increased, trustworthiness decreased. While across the two countries 68% of the second movers rewarded trust when a first mover would have earned only 0.1M, 51% were willing to do so when the first mover received 0.9M (almost full insurance). This pattern suggests that the second movers' behavior was affected by their social preferences. Either inequality aversion or reciprocity or both would lead them to reward trust less, the smaller was the payoff difference between themselves and the first mover, implying that the first mover was less vulnerable to betrayal. Note that efficiency preferences would have pushed in the opposite direction, leading second movers to reward trust more the less vulnerable was the first mover. Thus, our design allowed for the possibility that efficiency preferences would affect behavior, but such preferences would have yielded the opposite result from what we found.

For first movers, the patterns of behavior differed in the two countries. In Jordan, there was no significant relationship between the size of the betrayal payoff and willingness to trust. Jordanian first movers' behavior may have been inelastic to changes in the value of trusting because they did not respond to mitigating factors for cultural reasons, or because they anticipated their second movers' pattern of behavior. In the United States, first movers failed to anticipate the power of social preferences, and were more willing to trust the higher were their payoffs in case of betrayal. In fact, in our game, the institutional environment that would have maximized first movers' expected earnings given second movers' behavior in the Unites States would have been the least insured environment, the one offering the smallest betrayal payoff.

Calculating the elasticity of trust in the two countries, we find that, on average, a 10% increase in the first mover's betrayal payoff from its mean yields a 4% increase in the probability of trust in the United States, but does not cause a significant change in Jordan. Figure 7-1 plots the average trust levels across each of the insurance treatments for both Jordan and the United States. As reflected by the graph, more insurance on average yields higher rates of trust in the United States, but there is no clear relationship in Jordan.

Al-Ississ and Bohnet (2010) found significant cross-regional differences in the elasticity of trust, but not in the elasticity of trustworthiness. Thus, in addition to

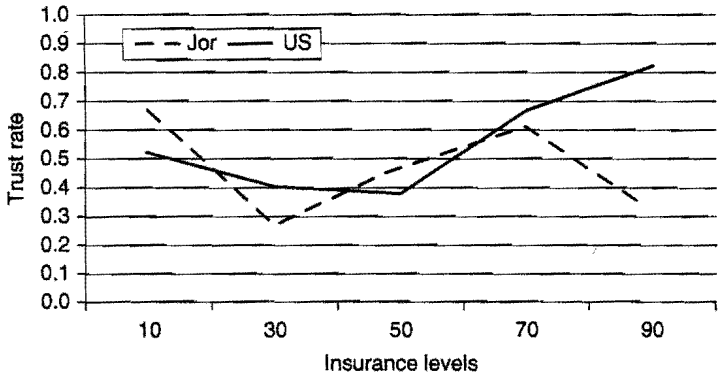

Figure 7-1 Trust Rates by Insurance Level in Jordan and the United States

the second mover's social preferences, cultural factors seem to matter. While we cannot exclude cross-regional differences in the ability to anticipate second mover behavior, with Jordanians more adept at this task than Americans, we suspect that the unfamiliarity and lack of acceptability of damages and insurance in Jordan contributed to the lack of trust responsiveness found there. In Islamic law, betrayal is normatively wrong and should not be encouraged by compensating the harmed parties for the losses incurred. Similarly, people should not be encouraged to trust lightly, as blind risk taking or speculation is discouraged; in its extreme it is incompatible with religious and legal doctrine (Vogel 1997).

■ TRUST RESPONSIVENESS TO CHANGES IN THE LIKELIHOOD OF TRUSTWORTHINESS

The likelihood of trustworthiness can be increased by placing a variety of incentives on the second mover. Rewards can make trustworthiness more valuable. The potential to punish a betrayal makes that action less attractive. We first examine how responsive trust is to changes in the likelihood of trustworthiness, independent of the way such changes are produced. We then focus on mechanisms that decrease the benefits of betrayal. Betrayers can be punished by third parties, such as the state in case of torts, or by the betrayed party itself. Bohnet et al. (2001) examined third-party punishment; here we focus on punishment by the betrayed party.

Changes in the Likelihood of Trustworthiness

We focus on people's responsiveness to changes in the likelihood of betrayal and compute the elasticity of trust. That elasticity describes how the percentage of those not trusting diminishes in response to a percentage reduction in those not trustworthy. Let t be the fraction of trusting principals, and w be the fraction of trustworthy agents. Our elasticity concept looks at the curve $t = f(w)$. The elasticity

measure at each point is thus [dt/(1 – t)]/[dw/(1 – w)]. This expression gives the percent reduction in those not trusting divided by the percent reduction in those not being trustworthy as trustworthiness increases. Since both numerator and denominator are positive, the elasticity will be positive.

To measure how many first movers are willing to trust for given fractions of trustworthy second movers, we employed an experimental design introduced by Bohnet and Zeckhauser (2004). It elicits first movers' minimum trustworthiness thresholds that would make them just willing to trust for a given set of payoffs in a binary-choice trust game. The method allows us to calculate how the fraction of trusting first movers responds to increases in the level of trustworthiness. Bohnet et al. (2010) measured the elasticity at each 10% increase of trustworthiness from 0% to 90%, thus looking only at decile intervals.[6] To get an overall elasticity measure, they averaged these nine numbers.

Trustworthiness thresholds were elicited by asking first movers to indicate how large the fraction of trustworthy second movers minimally would have to be for them to trust rather than to play safe. They informed first movers that they would be assumed to play safe (and thus earn the moderate payoff, M) if their trustworthiness threshold exceeded the actual likelihood of trustworthiness in the second-mover group. In contrast, if their trustworthiness threshold was lower than or equal to the actual level of trustworthiness, they would be assumed to trust and their payoff would depend on their specific second mover's choice. Second movers decided whether or not to reward trust in case it was offered to them at the same time that first movers chose their trustworthiness threshold. Thus, the strategy method was employed to conduct this game.

The more concerned a first mover is about trusting—because of her betrayal, risk, or inequality aversion—the higher will be her trustworthiness threshold. Those very averse to trusting could report a minimum threshold level of 100%; that is, they would be willing to offer trust only if every second mover was trustworthy.

Bohnet et al. (2008) proved that this mechanism is incentive compatible: a rational first mover should be indifferent between playing sure and trusting at her reported trustworthiness threshold. Bohnet et al. (2010) employed the binary-choice trust game using the threshold elicitation procedure described above in various countries in the Middle East and the West, also using parallel experiments to control for inequality and risk aversion. Here, we focus on a comparison in the elasticity of trust between the UAE and the United States. In the UAE, first movers requested a minimum trustworthiness threshold of 81% on average. The mode in the UAE was 1; for most Emirati first movers, absolute certainty was required to make them willing to trust. In the United States, first movers required an average trustworthiness level of 54% for them to be willing to trust. The cross-regional difference in trustworthiness thresholds was mainly caused by betrayal aversion

6. They excluded 100% as everyone is willing to trust if trustworthiness is guaranteed. Thus, the elasticity in the final decile interval is always 1.

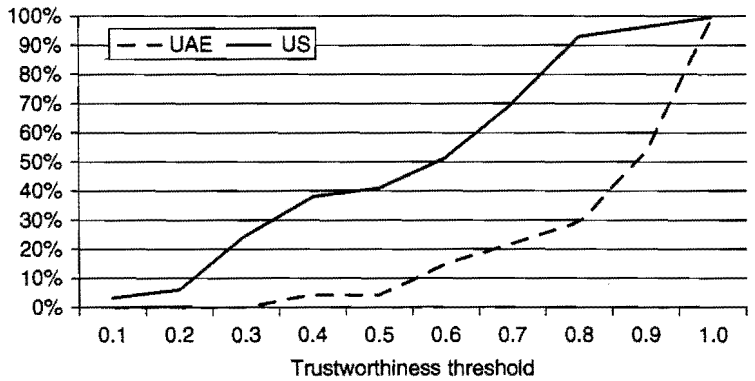

Figure 7-2 Cumulative Willingness to Trust in the UAE and the United States

being more pronounced in the UAE than in the United States, although Emiratis were also more risk averse than Americans.

As would be implied by stronger betrayal and risk aversion, trust was substantially less elastic to changes in the likelihood of trustworthiness in the UAE than in the United States. Figure 7-2 presents the elasticity of trust in response to the likelihood of trustworthiness—that is, the degree to which first movers trust more when second movers are more trustworthy. Using the elasticity measure introduced above, which is given by the percent reduction in those not trusting over the percent reduction in those not being trustworthy, the average trust elasticity is 0.21 in the UAE and 1.03 in the United States.

Punishment

Decreasing the benefit of betrayal by allowing betrayal to be punished affects the behavior of a money-maximizing second mover only if the punishment, or more precisely the expected punishment, is large enough to deter him from betraying. For other-regarding second movers who are concerned about inequality or reciprocity, the prediction is more complicated. If punishment costs the first mover less to impose than the second mover loses, it also diminishes the inequality between these two players in case of betrayal. In addition, and likely more importantly, the availability of punishment may also decrease the second mover's reciprocal inclination, as the extrinsic incentive of the punishment "crowds out" the intrinsic motivation to reward trust with trustworthiness (e.g., Bohnet et al., 2001; Fehr & Gächter, 2002). Thus, the net impact of punishment is unclear. The higher the likelihood and the size of the punishment, the more likely money-maximizing second movers will be deterred from betraying. Moreover, the mere presence of a punishment option may undermine the social motivation of those who would have rewarded trust voluntarily because they are inequality averse or reciprocally motivated.

Thus, how punishment affects trustworthiness is an empirical question. The relationship has been studied quite extensively in public goods games and

social dilemmas. (See Gächter & Herrmann, 2009, for a survey.) In individualistic societies, like most Western societies, costly punishment is mostly used by cooperators to reduce the income of free-riding group members (Herrmann et al., 2008). Such punishment of free-riders effectively increases cooperation.

In contrast to the punishment behavior observed in individualistic societies, in collectivistic societies punishment is also used by free-riders to punish cooperators. In particular, in Saudi Arabia, punishment of cooperators by free-riders occurred quite frequently. At the same time, cooperators were much less likely to punish free-riders in Saudi Arabia than in the United States. As a result of this "anti-social" punishment pattern, the cooperation rates in Saudi Arabia were lower in the presence of punishment than in its absence (Herrmann et al., 2008).

The effectiveness of endogenous punishment in dyadic trust relationships with sequential rather than simultaneous moves has received less study. In contrast to multiparty public goods games where even in a one-shot game, a punisher may feel like providing a public good to the community by punishing a free-rider (and paying the costs associated with such action), punishment in a trust game provides a purely private good and may be more appropriately referred to as "securing revenge."

We employed a standard binary-choice trust game and included a third stage where the first mover could decrease the second mover's payoffs in case the latter betrayed. We ran the experiments in Saudi Arabia and the United States.[7] The complete game structure was common knowledge and consisted of three sequential moves: the first mover's decision whether or not to trust, the second mover's decision whether or not to reward trust in case trust was offered to him, and then the first mover's decision whether or not to punish the second mover had he betrayed trust. In our game, the sure moderate payoff, M, was 10, the better payoff in case of trustworthiness, B, was 15, the first mover's low payoff in case of betrayal, L, was 8, and the second mover's high payoff in case of betrayal, H, was 22. The size of the punishment equaled P = 10 points, which was deducted from the second mover's betrayal payoff if the first mover chose to punish.

We kept incentives constant across regions, controlling for purchasing power parity. More generally, in all our experiments we followed the protocol for running experiments across regions by having the same experimenter run in the Middle East and the United States and translated and back-translated the instructions. In this specific experiment, cross-regional comparability was challenged by the fact that higher education is gender-segregated in Saudi Arabia, and thus replications in the United States had to involve both same-sex as well as mixed-sex treatments, as there is no obviously right comparison. We did not find any differences in behavior between our same-sex and mixed-sex treatments in the United States, and thus we aggregate the data below.

We held the punishment size constant but varied the first mover's cost of imposing punishment with a cost level of either 20% or 40% of the punishment size. As this was a one-shot interaction, first movers had no monetary incentive to punish,

7. The experiments in Saudi Arabia were made possible with the help of Khalid Al-Yahia.

and were they rational money-maximizers none would have punished. Of course, if revenge is sweet, and yields utility, punishment may well be taken. The second movers' optimization problem becomes more complicated as soon as we allow for first movers to be socially motivated and, for example, care about reciprocity. Second movers have to be concerned that first movers will find the benefits from revenge outweigh the costs they have to bear in order to punish.

The control treatment and the two punishment treatments were run with 192 study participants in the Harvard Decision Science Laboratory in the United States and with 190 participants at Imam Muhammad bin Saud University in Riyadh, Saudi Arabia. Table 7-1 presents summary results for each treatment and country.

The availability of punishment increased willingness to trust in Saudi Arabia but not in the United States. It did not affect willingness to reward trust in either of the two countries. Willingness to punish seemed to increase as the cost of punishment decreased, but the difference in punishment rates by cost level was significant only in Saudi Arabia. Comparing across the two countries, Americans were more likely to trust than Saudis in the control treatment without a punishment option and generally were more likely to reward trust. When comparing treatment by treatment, however, that difference was significant only for a cost level of 4. The differences in punishment rates were not significant, as our punishment sample size was small at that point.

Across the two countries, the punishment option did not deter second movers from betraying. Based on our payoffs, assuming they were known, punishment rates would have had to be 75% to deter money-maximizing second movers from betraying. A post-experimental questionnaire showed that for both cost levels, second movers expected only about 50% of the first movers to actually punish given betrayal. In addition, trustworthiness rates may not have been higher in the punishment treatments than in the control treatment as the punishment possibility may have crowded out the socially oriented second movers' willingness to reward trust voluntarily. This interpretation of our findings is in line with our earlier work on the effectiveness of exogenous punishment probabilistically imposed by a third party, such as the state (Bohnet et al., 2001).

Moreover, if punishment is not high enough to substantially deter second movers from betraying, its availability may hurt first movers through another route. In some sense, it frames the trust game much more as a mere expected value exercise. When threats hover, the spirit of reciprocity may get diminished

TABLE 7-1 *Punishment Treatment Results for Saudi Arabia and the United States*

Country	Treatment		Number of Participants	Fraction of 1st Movers Who Trusted	Fraction of 2nd Movers Who Rewarded Trust	Fraction of 1st Movers Who Punished Betrayal
Saudi Arabia	Control		72	47%	53%	
	Revenge	Cost = 2	32	69%	55%	100%
		Cost = 4	86	67%	45%	6%
United States	Control		60	77%	65%	
	Revenge	Cost = 2	56	57%	63%	50%
		Cost = 4	76	66%	72%	29%

attention. Similar negative effects from the availability of sanctions on cooperation and trustworthiness have been found by Tenbrunsel and Messick (1999), Gneezy and Rustichini (2000), Fehr and Gächter (2002), and Fehr and List (2004) and more generally in contract enforcement by Malhotra and Murnighan (2002). Sitkin and Roth (1993, p. 376) observe that "legalistic remedies can erode the interpersonal foundations of a relationship they are intended to bolster because they replace reliance on an individual's goodwill with objective, formal requirements."

Interestingly, trust responded positively to the punishment option in Saudi Arabia, although the increased trust rate was not warranted by an increased trustworthiness rate. As the Saudis were not more likely to punish than Americans across both cost levels, cross-regional differences in the utility of taking revenge do not seem to be able to account for the differential responsiveness of trust to a punishment option.

While our sample is too small to draw any definite conclusions, we observe that willingness to punish seems to be more sensitive to the cost level in Saudi Arabia than in the United States. This observation is consistent with earlier work on willingness to punish in public goods games, which found that the use of punishment in Saudi Arabia was much more sensitive to the pattern of final *relative* outcomes than it was in the United States.[8]

■ DISCUSSION AND CONCLUSIONS

Mechanisms aimed at mitigating the cost of betrayal, such as damages or insurance provision, seem to work better in the United States, and arrangements focusing on preventing the occurrence of betrayal, such as a punishment threat, have greater impact in the three countries of interest in the Arab Middle East. In our experiments, trust was promoted by decreasing the cost of betrayal in the United States but not in Jordan. Punishment functioned differently. Giving the first mover the option to take revenge at a price should she be betrayed enhanced trust in Saudi Arabia but not in the United States.[9]

We take this as a first indication that the elasticity of trust vis-à-vis a given intervention depends on customary practices, which provide reference points for expectations. In the Arab Middle East, people are used to trust being promoted through the prevention of betrayal by operating in close-knit groups, such as clans or tribes. In such groups, games are repeated and reputations play a prominent role. In fact, when we asked for people's trustworthiness reference points in the

8. In our game, paying a cost of 4 points to punish a betraying second mover slightly worsened the relative position of the first mover. It resulted in a final payoff of 4 points for the first and 12 points for the second mover, or a 1:3 payoff ratio, as compared to a 1:2.75 payoff ratio without the enforcement of punishment. In contrast, when punishment cost only 2 points, imposing punishment improved the relative standing of the first mover. It produced a 1:2 payoff ratio as compared to a 1:2.75 ratio without the enforcement of punishment. Controlling for purchasing power parity, at the time of the study points could be converted 1:1 into dollars.

9. We follow our gender convention for labeling first movers here even though there were no female participants in Saudi Arabia.

UAE, the modal response was that they would only trust if everyone was trustworthy. In contrast, in the United States, people were willing to trust for much lower levels of trustworthiness in an otherwise identical situation. Americans are used to trusting strangers and experiencing betrayal that then would be compensated by damages and other legal remedies.

However, neither the mitigation- nor the prevention-focused approach offers a sustainable solution. People's trust was not warranted by the levels of trustworthiness experienced. While trust increased as the cost of being betrayed decreased in the United States, a money-maximizing first mover would have preferred an environment offering her less insurance to one offering her more. Trustworthiness increased, the smaller were betrayal payoffs, leading to a higher expected value of trusting when the first mover's costs of betrayal were high than when they were low.

Similarly, second movers' behavior hardly responded to a punishment threat, possibly because the expected punishment was too low to deter money-maximizing second movers from betraying or because it undermined the intrinsic motivation of those second movers who would have honored trust voluntarily absent this instrument. Thus, social preferences, which in the first place make trustworthiness possible, can undermine the effectiveness of an intervention aimed at decreasing the material cost or the material benefits of betrayal.

For both interventions, first movers responded in a manner consistent with traditional views of their culture but did not anticipate the interaction between a given intervention and second movers' social preferences. While the elasticity of trust seems to be strongly related to the situational context, the elasticity of trustworthiness is much more similar across the two regions studied and influenced by basic human motivations, such as fairness and reciprocity. This is in line with the general finding that stereotypes about typical behaviors more strongly affect people's beliefs than actual behavior (Croson & Gneezy, 2009). Deciding whether or not to trust a stranger, which in our case one only knows that this person is a fellow participant, is likely based on beliefs about the group the fellow participants belong to—Americans, Emiratis, Jordanians, or Saudis. Deciding whether or not to be trustworthy does not involve beliefs but does depend on one's intrinsic motivation.

Our data shed light on how trust might best be produced in the two different contexts. Trust can be promoted by mitigation of the cost in the United States (e.g., through damages or insurance provisions) but much less so in the Arab Middle East. To foster trust in the Middle East, preventing the occurrence of betrayal must be the goal, through mechanisms affecting the trusted party's incentives to honor trust, such as repeated interactions, reputational concerns, or the threat of punishment. We expect cross-cultural misunderstandings and difficulties in negotiations if the trust-production processes parties are accustomed to are not taken into account. Clearly, however, any cross-cultural interaction is complicated by the question of who adjusts to whom and of whether a third, supra-cultural solution or compromise might be possible. Our studies primarily speak to how trust can be promoted within a specific context, offering some food for thought for cross-cultural interactions.

■ REFERENCES

Al-Ississ, M., & Bohnet, I. (2011). *Does insurance increase trust? Experimental evidence on institutional design.* Working paper.

Al-Suwailem, S. (2000a). *Decision under uncertainty: An Islamic perspective.* Working paper, Al-Rajhi Banking and Investment Corporation, Riyadh, Saudi Arabia.

Al-Suwailem, S. (2000b). Towards an objective measure of gharar in exchange. *Islamic Economic Studies, 7,* 61–102.

Arab Human Development Report 2002, United Nations Development Program.

Arab Human Development Report 2004, United Nations Development Program.

Baroudi, S. E. (2004). The 2002 Arab Human Development Report: Implications for democracy. *Middle East Policy, 11,* 132–141.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes, 63,* 131–144.

Bohnet, I., Frey, B. S., & Huck, S. (2001). More order with less law: On contract enforcement, trust, and crowding. *American Political Science Review, 95*(1), 131–144.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior and Organization, 55,* 467–484.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal aversion. *American Economic Review, 98,* 294–310.

Bohnet, I., Herrmann, B., & Zeckhauser, R. (2010). Trust and the Reference Points for Trustworthiness in Gulf and Western Countries. *Quarterly Journal of Economics, 125*(2), 811–828.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90*(1), 166–193.

Camerer, C., & Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica, 56,* 1–36.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature, 47*(2), 448–474.

Dasgupta, P. (1988). Trust as commodity. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 49–72). Oxford: Blackwell.

Economist, "A long way to go," April 9, 2005, pp. 36–38.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415,* 137–140.

Fehr, E., & List, J. (2004). The hidden costs and returns of incentives: Trust and trustworthiness among CEOs. *Journal of the European Economic Association, 2*(5), 743–771.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*(3), 817–868.

Fehr, E., & Schmidt, K. (2002). Theories of fairness and reciprocity: Evidence and economic applications." In M. Dewatripont & S. Turnovsky (Eds.), *Advances in Economics and Econometrics: 8th World Congress, Econometric Society Monographs* (pp. 208–257). Cambridge, MA: Cambridge University Press.

Gächter, S., & Herrmann, B. (2009). Reciprocity, culture, and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B—Biological Sciences, 364,* 791–806.

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies, 29,* 1–17.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The Maghribi trader's coalition. *American Economic Review, 83,* 525–548.

Greif, A. (1994). Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *Journal of Political Economy, 102,* 912–950.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*, 1480–1484.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science, 319*, 1362–1367.

Hofstede, G. (2001). *Culture's consequences* (2nd ed.). Thousand Oaks, CA: Sage.

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology, 133*, 23–30.

Inglehart, R. (2005). The world views of Islamic publics in global perspective. In M. Moaddel (Ed.), *Worldviews of Islamic publics*. New York: Palgrave.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47*, 263–291.

Köszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics, 121*, 1133–1165.

Kreps, D. (1990). Corporate culture and economic theory. In J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on positive political economy* (pp. 90–143). Cambridge, MA: Cambridge University Press.

Kuran, T. (1995). Islamic economics and the Islamic subeconomy. *Journal of Economic Perspectives, 9*, 155–173.

Kuran, T. (2004). *Islam and mammon*. Princeton, NJ: Princeton University Press.

Kuran, T. (2005). The logic of financial westernization in the Middle East. *Journal of Economic Behavior and Organization, 56*, 593–616.

Layne, L. L. (1987). "Tribalism": National representations of tribal life in Jordan. *Urban Anthropology and Studies of Cultural Systems and World Economic Development, 16*(2), 183–203.

Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology, 57*(3), 426–441.

Malhotra, D., & Murnighan, J. K. (2002). The effects of formal and informal contracts on interpersonal trust. *Administrative Science Quarterly, 47*, 534–559.

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger and spite: emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes, 68*, 208–224.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83*, 1281–1302.

Rosen, L. (2000). *The justice of Islam*. Oxford: Oxford University Press.

Rutte, C. G., Wilke, H. A. M., & Messick, D. M. (1987). Scarcity or abundance caused by people or the environment as determinants of behavior in the resource dilemma. *Journal of Experimental Social Psychology, 23*, 208–216.

Sharabi, H. (1988). *Neopatriarchy*. New York: Oxford University Press.

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science, 4*, 367–392.

Tenbrunsel, A. E., & Messick, D. M. (1999). Sanction systems, decision frames and cooperation. *Administrative Science Quarterly, 44*, 684–707.

Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.

Vogel, F. E. (1997). *The contract law of Islam and of the Arab Middle East*. Working Paper, Harvard Law School.

World Values Survey (2003). See http://www.worldvaluessurvey.org

APPENDIX 7-1

TABLE 7-A1 *Questionnaire*[10]

1. Imagine that you are confronted with 100 people from one of the following groups. You do not know anything else but the information provided below about these groups. Please indicate in each case how many out of the 100 people you think would be worthy of your trust. Assume that the people in these groups do not know anything about you (i.e., they do not know your gender, your nationality, ethnicity or religious affiliation, etc.).
   a. Men: I would trust _____ out of 100 men.
   b. Women: I would trust _____ out of 100 women.
   c. Africans: I would trust _____ out of 100 Africans.
   d. Americans (USA): I would trust _____ out of 100 Americans.
   e. Asians: I would trust _____ out of 100 Asians.
   f. Europeans: I would trust _____ out of 100 Europeans.
   g. Middle East: I would trust _____ out of 100 Middle Easterners.
   h. South America: I would trust _____ out of 100 South Americans.
   i. Buddhists: I would trust _____ out of 100 Buddhists.
   j. Christians: I would trust _____ out of 100 Christians.
   k. Hindus: I would trust _____ out of 100 Hindus.
   l. Jews: I would trust _____ out of 100 Jews.
   m. Muslims: I would trust _____ out of 100 Muslims.
   n. Non-religious: I would trust _____ out of 100 non-religious people.

10. Questions h and n were not included in United Arab Emirates. Pilots suggested that they are either not meaningful or perceived as offensive. Rather, a question on trust in Emiratis was added.